

Task 5 - Natural Language Prompts for LLM

This document includes eight natural language prompts designed to be answered by a Large Language Model (LLM) based on descriptive statistics calculated from the 2025 Syracuse University Women's Lacrosse dataset. These prompts were crafted to balance simple factual retrieval (e.g., totals, averages) with higher-level reasoning (e.g., identifying the most improved player). The prompts were selected based on the dataset's available fields and aligned with questions a coach or analyst might realistically ask.

Prompt List:

1. How many games did this team play?
2. What was the team's average goals per game?
3. What was the team's average assists per game?
4. Which player had the highest shot accuracy?
5. Which player committed the most turnovers?
6. Which player had the most ground balls?
7. Who scored the most goals this season?
8. Who was the most improved player this season, and why?

Why These Prompts Were Chosen:

These prompts were selected to cover a range of common statistical evaluations:

- Questions 1 to 3 target basic team-level performance metrics.
- Questions 4 to 7 analyze individual contributions (accuracy, scoring, possession).
- Question 8 introduces a more complex scenario requiring custom metric creation due to lack of per-game or season-over-season data.

This mix provides a strong base for testing both factual correctness and reasoning abilities in LLMs.